

#InternetArchive : 487 billones de documentos y páginas de Internet

Carolina De Volder

carolina_devolder@yahoo.com.ar

[Internet Archive](https://archive.org/) -de ahora en más IA- es un archivo o biblioteca digital gestionada por una organización sin fines de lucro dedicada a la preservación de archivos, capturas de sitios públicos de la web, recursos multimedia y software. Fue creada en 1996 y existe gracias al apoyo de Alexa Internet y de otros colaboradores que han donado materiales y colecciones como la Biblioteca del Congreso de los Estados Unidos, la Biblioteca Nacional de España, JSTOR, Arxiv y otras tantas bibliotecas y repositorios del mundo.

Alberga una gran cantidad de archivos de muchos tipos como audio, video y texto, muchos de ellos endominio público, o con licencias de tipo Creative Commons u otras licencias que permiten su distribución gratuita. Pero lo más interesante, es que conserva todo lo que se produce en la red, aunque haya sido borrado o modificado y ya no se recupere a través de los buscadores comunes.

En IA se recopilan un billón de páginas web cada semana. La razón de este afán de conservación es que las creaciones y sitios web en la red son muy poco estables, cambian y desaparecen con mucha frecuencia. Se calcula que la vida media de un contenido web es de 100 días, tal y como comenta Brewster Kahle, el fundador de Internet Archive.



Fuente: <https://archive.org/>

Tipos de colecciones

IA ofrece acceso a 7 tipos de colecciones o secciones diferentes según tipo de material: web, texto, video, audio, imágenes, software, conciertos y colecciones.

La sección web se explicará más adelante (ver WayBackMachine).

La sección de textos alberga libros digitalizados y colecciones de varias bibliotecas, instituciones y fundaciones culturales procedentes de todo el mundo. En la actualidad ofrece más de 10 millones de libros totalmente accesibles y textos. También hay una colección de más de 300.000 libros electrónicos que pueden ser consultados o descargados en [OpenLibrary.org](https://openlibrary.org) (ver más abajo) .

La sección de videos permite descargar o ver películas y vídeos gratis. Contiene películas digitales subidas por los usuarios de los archivos que van desde largometrajes clásicos, a las emisiones de noticias diarias, dibujos animados y conciertos. Muchos de estos videos están disponibles para su descarga gratuita. En la actualidad posee 2,537,667 objetos diferentes de 4,225 diferentes colecciones.

La sección de audios incluye música, audiolibros, grabaciones de programas de radio antiguos y noticieros, entre otros archivos sonoros. Hay casi 3 millones de grabaciones de 9,182 colecciones diferentes.

La sección de imágenes contiene imágenes digitales subidas por los usuarios del archivo que van desde mapas a imágenes astronómicas o fotografías de obras de arte. Muchas de estas imágenes están disponibles para su descarga gratuita.. Contiene por ejemplo colecciones de imágenes del Metropolitan Museum of Art, mapas del Servicio Geológico de los Estados Unidos, una colección de la NASA, una colección de 500.000 carátulas de discos, entre otras. En la actualidad posee 1,168,657 objetos de 140 colecciones diferentes.

IA se asoció con etree.org -una comunidad comprometida con la prestación de los conciertos en directo- para preservar y archivar tantos conciertos en directo como sea posible para que puedan disfrutar las actuales y futuras generaciones. Toda la música en esta colección es estrictamente no comercial, tanto para el acceso como para la distribución.

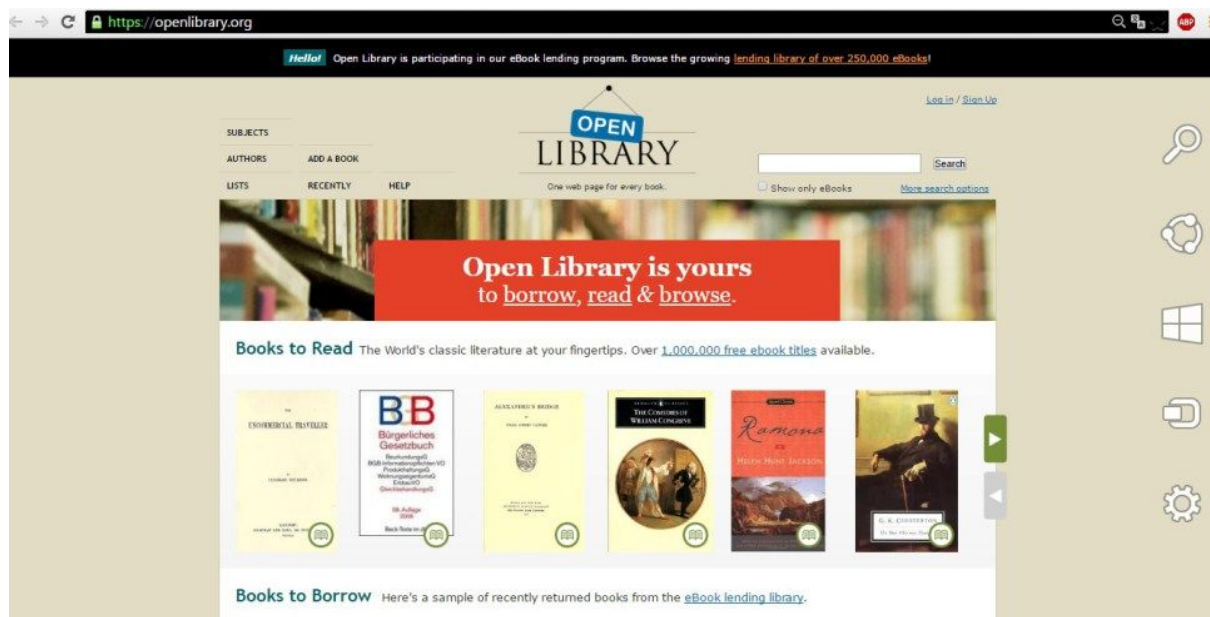
La sección de software es alberga la colección de software más grande del mundo, proporcionando acceso instantáneo a millones de programas, imágenes de CD-ROM, documentación y multimedia.

Otros productos

Open Library

Es una biblioteca digital, libre, que ofrece acceso a de más de 1 millón de eBooks que se pueden leer o descargar para su lectura fuera de línea. Fue inicialmente liderado por Aaron Swartz. El 16 de julio de sus veintiún años, Aaron escribió en su blog:

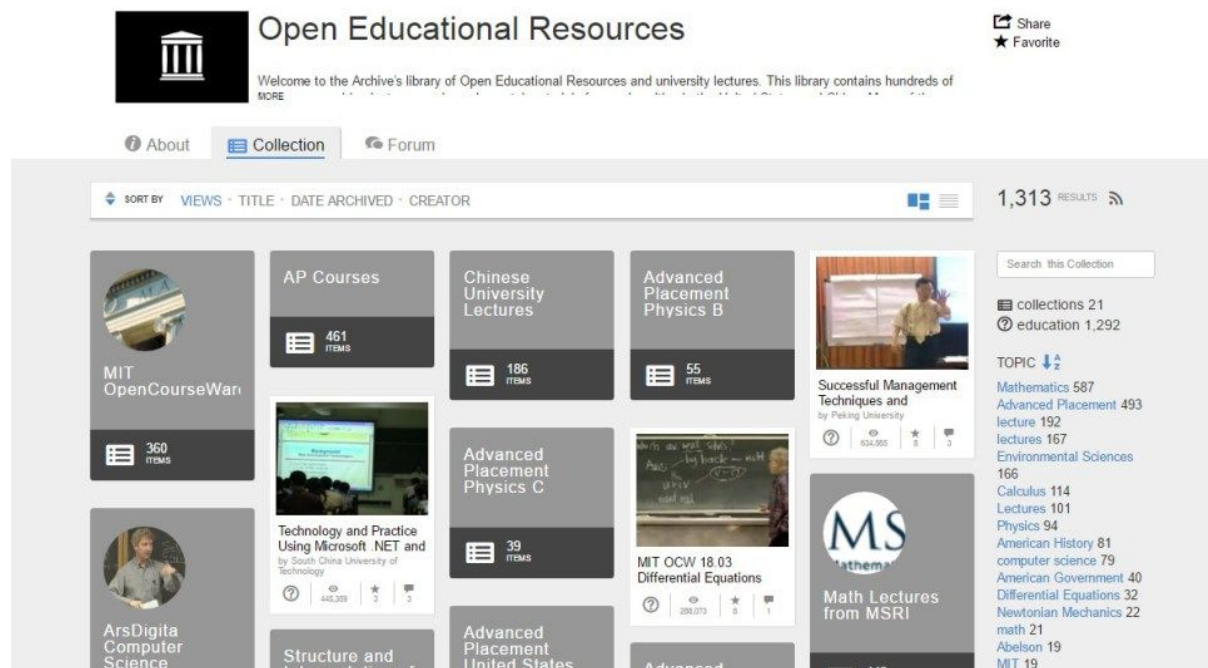
Nuestro objetivo es la construcción de la biblioteca más grande del mundo, para luego ponerla en internet de forma libre/gratuita; y que todos la usen y editen. Los libros son el lugar al que vas cuando tienes algo que quieras compartir con el mundo —son el legado cultural de nuestro planeta. Y nunca ha habido un intento más grande para juntarlos a todos.



Fuente: OpenLibrary <https://openlibrary.org/>

Education

Es una biblioteca de recursos de educación abierta que contiene cientos de cursos gratuitos, conferencias, videos y materiales complementarios de universidades de Estados Unidos y China.



Fuente: Open education <https://archive.org/details/education>

NASA Images

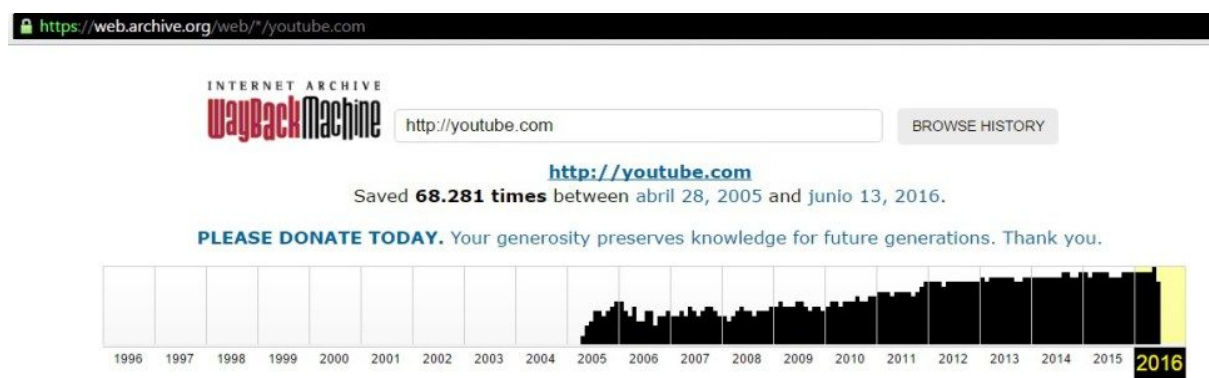
Fue creada en colaboración entre [la NASA](#) y el AI para proporcionar acceso público a las colecciones de imágenes, video y audio de la NASA. Es la colección más grande de recursos de la NASA disponibles desde un solo sitio. Uno de sus objetivos es promover la educación y facilitar la enseñanza de las matemáticas y las ciencias en todos los niveles, y para despertar el interés general y entusiasmo en torno a la exploración espacial, aeronáutica, y la astronomía.

Bookmobile

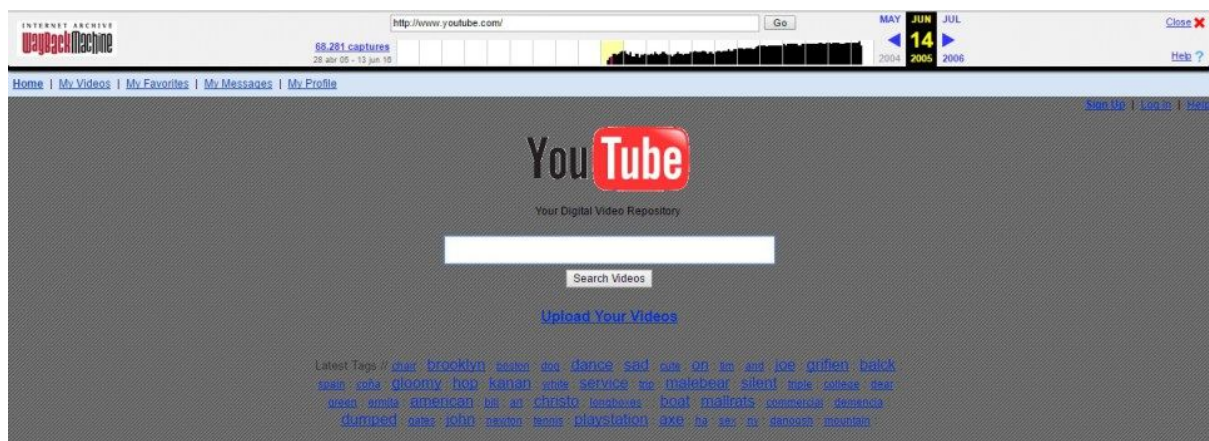
Es una biblioteca digital móvil capaz de descargar libros de dominio público a través de Internet a través de satélite e imprimirlas en cualquier momento y en cualquier lugar, para cualquier persona. El proyecto consiste en llevar a lugares de Estados Unidos libros de dominio público dentro de una camioneta, pero no van impresos, sino que por el contrario llevan las impresoras, máquinas para encuadernar y dispositivos para conectarse a internet y descargar los libros. Estas herramientas les permiten imprimir, ensamblar y encuadernar el libro en frente de la persona que lo solicite, en muy corto tiempo y por un valor entre \$1 y \$2 dólares.

La máquina del tiempo: [Wayback Machine](#)

Es el archivo web de IA, lanzado en 1996, que contiene más de 487 billones de capturas web en más de 40 idiomas. Lo “curioso” de este proyecto es justamente que permite consultar consultar un sitio web o página a través de la historia. Wayback Machine funciona de una manera simple: basta teclear la dirección url de una página web para ver cuál es la última copia que hay guardada en el archivo. Si queremos ver cómo era hace un tiempo una página en cuestión, entonces Wayback machine nos pedirá en qué fecha y año queremos visitar la página. También es posible ver de forma gráfica esa historia a través de una barra en la parte superior de la pantalla. La longitud de las barras del gráfico indica en qué meses se hicieron más copias.



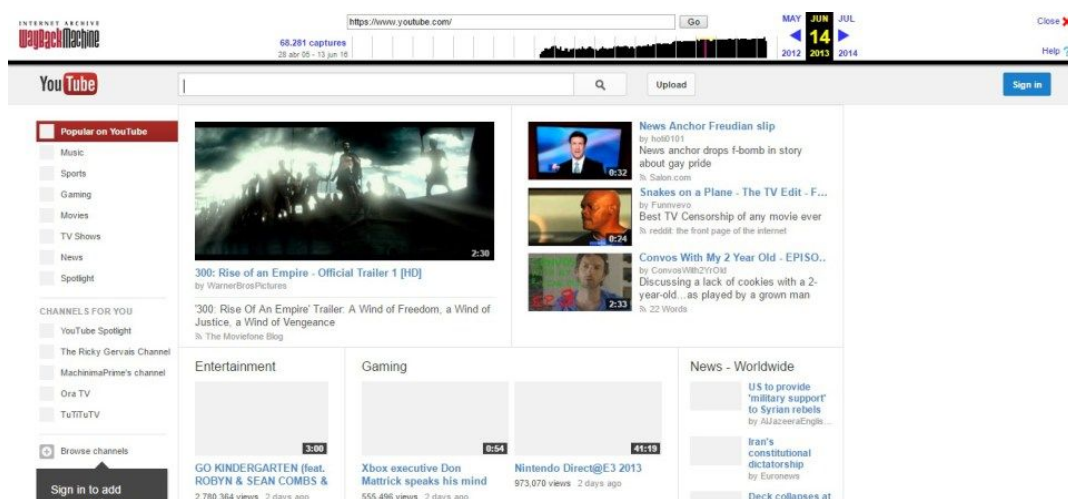
Youtube en WayBackMachine



Youtube en WayBackMachine - 14 de junio de 2005



Youtube en WayBackMachine - 14 de junio de 2009



Youtube en WayBackMachine - 14 de junio de 2013

Internet Archive en cifras

- 10 petabytes. Cantidad de información almacenada en IA. Un petabyte equivale a un millón de gigabytes o a 10 bytes elevados a la 15ª potencia. Los 10 petabytes de información que almacena el IA equivalen a 10 millones de gigabytes, más o menos lo mismo que 10.000 millones de libros -66 veces más que la Biblioteca del Congreso de los EE UU, la mayor del mundo (150 millones de ejemplares) y 11.000 veces más que la mítica de Alejandría, que albergaba 900.000 manuscritos.
- 10.000 millones de libros impresos -400 veces más que los almacenados en la Biblioteca Nacional de España- serían necesarios para contener la misma información.
- 487 billones de páginas web. La Wayback Machine del IA ha rastreado y guardado todas las webs de Internet.
- Más de 2 millones de archivos de audio. El IA también digitaliza y ofrece gratis música, informativos de televisión, películas y libros electrónicos.

Cómo las bibliotecas pueden usar Internet Archive

Además de los servicios o proyectos antes mencionados, en IA también es posible que individuos utilicen el servicio de manera autónoma para publicar o depositar sus propios archivos de texto, audio, imagen y video. Una de las cosas más interesantes de esta plataforma es que no se especializa en un único formato, sino que se pueden subir múltiples medios en un sólo lugar.

Hay que tener en cuenta que subir una obra a IA es hacerla pública, brindando acceso para todo el mundo. No es un material que está en la nube para almacenar archivos personales, como Dropbox o Google Drive. Tampoco es una plataforma de venta de ebooks como Amazon o Google Play Books, ni un servicio online comercial de almacenamiento como Issuu o Scribd.

Al subir un libro por ejemplo a IA tenemos la seguridad de que no le cedemos a nadie los derechos, que no habrá requisitos ni restricciones para los lectores (pueden acceder sin necesidad de crearse una cuenta), ni se hará una explotación publicitaria de las obras. Es posible subir todo tipo de material y el servicio se mantendrá gratuito, al menos mientras exista IA. Es por estos motivos que puede ser una muy buena solución para las bibliotecas que quieran hacer accesibles sus obras y colecciones.

Mini instructivo para subir un libro a IA

Antes de subir material a IA hay que evaluar si el material es de dominio público o no. En general hay dos posibilidades:

1. Si eres autora o autor del libro o representas a alguna entidad que cuenta con los
2. derechos patrimoniales, puedes licenciarlo con licencias Creative Commons.
3. Si trabajas en una Bibliotecas o archivo puedes identificar acervos que se pueden compartir por estar ya licenciados o en dominio público.

Antes de subir el documento es necesario preparar el PDF, teniendo en cuenta dos aspectos fundamentales:

- Que el libro tenga metadatos básicos: título, autor, fecha de publicación, etiquetas y descripciones.
- Que incluya algunos marcadores internos para la navegación básica dentro del archivo.

Generalmente, los procesadores de texto como LibreOffice Writer y MS Word, así como otros programas que se usan para crear documentos PDF, permiten editar las Propiedades de un

documento, donde podemos especificar algunos datos básicos. Asimismo, la mayoría de los procesadores de texto permiten crear una tabla de contenidos o índice

automáticamente si se aplican los estilos de títulos y subtítulos. También podemos crear manualmente marcadores para establecer hipervínculos internos. Al finalizar y exportar a PDF, estos datos y marcadores también se exportan y el libro queda listo para la circulación online en condiciones óptimas. Otra opción es utilizar un software de edición de PDF. No hace falta uno muy completo ni complejo, basta con que permita editar los metadatos y crear tablas de contenido con marcadores.

Luego ya podemos subir la obra a IA. Para eso accedemos al [sitio](#) y nos logueamos (sacar una cuenta es muy sencillo: solo hay que proveer un correo electrónico y contraseña). Aunque también es posible subir material sin loguearse, haciendo clic en Register for free. Para cargar un nuevo archivo, ir al ícono de la flecha apuntando hacia arriba.

En la siguiente pantalla, hacer clic en Upload Files. Arrastrar el archivo desde la ubicación en nuestro equipo. Es posible cargar varios archivos para el mismo ítem. Por ejemplo: versión en PDF, TXT y EPUB, si las tuviéramos. A continuación completamos la ficha con los datos básicos del libro (título, autor, resumen, licencia, etc.). Si queremos podemos modificar la url, ya que por defecto tomará el nombre del título de la obra. En el campo Collection elegimos Community Texts.

También es posible crear una colección propia. Para eso hay que solicitarlo por mail a: collections-service@archive.org

Una vez completos estos datos, confirmamos con el botón Upload and Create Your Item. Una vez cargado, IA tardará unos minutos en terminar el procesamiento para que el ítem se previsualice. Sin embargo, ya está disponible para que los usuarios lo descarguen en distintos formatos. Finalmente, cuando termine el procesamiento del ítem, veremos una atractiva visualización que emula el pasar de páginas de un libro. En esta visualización podemos hacer zoom para alejar y acercar las páginas y realizar búsquedas a texto completo dentro de la obra (siempre que los caracteres sean reconocidos por el software de Internet Archive). Cuando se encuentran las palabras buscadas, aparecen en la parte de abajo unos marcadores. Si nos posicionamos encima, se despliega un fragmento con el contexto de la palabra o frase buscada. Pinchando en el marcador nos trasladamos hasta la ubicación específica en el texto.

Con el ícono Share this item podemos elegir entre varias opciones de difusión del libro, compartir en distintas redes sociales o copiar el código para insertar el libro y que sea visualizado en una página web externa.

Aquí una captura de pantalla de un documento que subí.



Fuente: https://archive.org/details/Cita_documental

Fuentes consultadas

- Álvarez de Toledo, Ma Luisa (2015) Internet Archive. Por qué es importante. Para qué sirve <http://www.infobiblio.es/internet-archive-por-que-es-importante-para-que-sirve/>
- Delgado, Andrés (2015). La biblioteca abierta - OpenLibrary.org. <http://andres.delgado.ec/2015/08/29/la-biblioteca-abierta-openlibrary-org/>
- Fossatti, Mariana. Tutorial: ¿Cómo subir libros a Internet Archive? (2016). Material del curso online de digitalización y comunicación de archivos culturales (Ártica). <http://www.articaonline.com/2016/06/tutorial-compartir-libros-en-internet-archivo/>
- González, José Ángel (2016). Internet Archive: la biblioteca de Alejandría del siglo XXI. <http://www.20minutos.es/noticia/1927160/0/internet-archive/historia/red/>
- Internet Archive. <https://archive.org/>
- Internet Archive en Wikipedia. https://es.wikipedia.org/wiki/Internet_Archive
- Wayback Machine en Wikipedia. https://es.wikipedia.org/wiki/Wayback_Machine

Post publicado en Infotecarios (<http://www.infotecarios.com/internet-archive/>) el 13/06/16